

Auto Scaling Getting Started Product Documentation





Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice

🔗 Tencent Cloud

All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.



Contents

Getting Started

Creating a Scaling Plan in 5 Minutes

Step 1:Creating a Launch Configuration

Step 2:Creating a Scaling Group

Step 3:Creating a Scaling Policy

Getting Started Creating a Scaling Plan in 5 Minutes

Last updated : 2024-01-08 17:53:29

Feature Overview

This document describes how to create a complete AS scheme in three steps.



Note:

This document describes how to create a scheme in the AS console. If you prefer to use APIs, follow the instructions in Introduction.

Directions

Create a complete AS scheme as instructed in the following documents:

Step 1: Creating a Launch Configuration

Step 2: Creating a Scaling Group

Step 3: Creating a Scaling Policy

Step 1:Creating a Launch Configuration

Last updated : 2024-01-08 17:53:29

Overview

A launch configuration defines the configuration information of CVM instances used for auto scaling, including their images, storage, networks, security groups, login methods, and other configuration information. **Note:**

Creating a launch configuration is free of charge.

Directions

Selecting a region

- 1. Log in to the Auto Scaling console and click Launch Configuration in the left sidebar.
- 2. At the top of the **Launch configuration** page, select the project and region of the launch configuration.

Auto Scaling	Launch configuration All projects V Guangzhou V
Scaling group	
타 Launch	 Launch Configuration now supports multiple models, allowing you to choose multiple similar models to imp
conngulation	Create

CVM instances and CLB instances must be in the same region as the one specified for launch configuration. For example, if the Guangzhou region is specified for the launch configuration, only CVM instances in Guangzhou will be automatically added to the scaling group. For a scaling group in Guangzhou, you cannot add CVM instances or bind CLB instances from other regions (such as Shanghai, Beijing, Hong Kong (China), or Toronto). 3. Click **Create** to go to the **Create a launch configuration** page.

Selecting a model

Set up the launch configuration name, availability zone, and model.

1.Select mo	2.	Complete coni	Iguration	3.Con	firm co	ntiguration			
Launch configuration name			You	u can enter 60 o	haracters	5			
Billing mode	Pay as you go	Spot instance	es Detailed	comparison e					
Region	Guangzhou								
Availability zone	All AZs	Guangzhou Zon	e 2 Guan	gzhou Zone 3	Guan	gzhou Zone 4	4 Guangzho	ou Zone 5	Guangzh
	Guangzhou Zo The information of A	ne 7 Promo Z is not included in th	e launch configu	uration. The AZ se	lected here	e is only used to	list available insta	ince types in the s	selected AJ
Instance	All CPUs	v	All Mems	v					
	All models	Standard	High IO	MEM-optimiz	zed	Compute	GPU-based	Big Data	Clo
	All types	Standard S6	Standard SA	3 NEW Stand	lard S5	Standard	SA2 Standa	ard S4 Star	ndard S3
	Standard Netw	ork-optimized SN?	ne Stand	lard SA1 S	tandard N	letwork-optim	nized S2ne	Standard S2	Stand

Launch configuration name: Set the name of the launch configuration.

Billing mode: Support Pay As You Go and Spot Instance.

Availability zone, model: Select the model of the instance to be bound with the scaling group.

Selecting images, storage, and bandwidth

1. When creating a launch configuration, you can use a public image, custom image or shared image. For more information, see Image Overview.

Image	Public image	Custom image	Shared image			
	TencentOS	~ 6	i4-bit	÷	Please select	v

We recommend that you use a custom image where the application environment has already been deployed for the following reasons:

If you select a **public image**, the CVM instances created in a scaling group will have an operating system without the application environment. Then, you need to manually deploy the application environment.

If you select a **custom image**, you can use the image created for a CVM instance with an environment that has been deployed to batch create CVM instances have the same software environment as the original CVM, so as to



implement batch deployment.

Note:

For more information about creating images for CVM instance to be bound to a scaling group, see Creating a Custom Image.

2. Set the disks in the launch configuration.

System disk	Premium cloud storage 🖌 - 50 + GB Learn r	tore #
Data disk	System disk type cannot be changed after purchase Create cloud data disk You can add 20 data disk(s)	

If you specify a cloud disk as the system disk, you can create a data disk using data disk snapshots:

Users with a large amount of data often use data disks to store data. You can create a snapshot for data disk A, and use this snapshot to quickly clone multiple disks for rapid server deployment.

When a new CVM instance is automatically added for Auto Scaling, if you've specified a snapshot for the data disk in the launch configuration, CBS automatically mounts a data disk to the launched CVM instance to copy data. If a data disk snapshot is specified in the launch configuration, ensure that the data disk can be automatically mounted correctly so that the scaling group can be scaled out automatically. The snapshot of the data disk of the original instance should be taken before auto scaling is configured so that data disks can be automatically mounted when new CVM instances are activated. For more information, see Attaching Cloud Disks.

3. An independent public IP address is allocated by default at no charge. Please select a network billing method based on your actual needs.

ublic network	Assign a ded	icated public IP for free					
andwidth	By traffic	Bandwidth package	Detailed comparison ⊠				
		11	I	Т	-	5	Mbps
1	Mbos	5Mbps	20Mbps	100Mbps			

Note:

Auto Scaling is free of charge, but the added CVM instances, disks, and networks are billed in pay-as-you-go mode. Prices will be shown based on your configurations.

Setting information

1. In the **Configure the CVM** step, select the login method and security group. By default, CVM instances created by Auto Scaling are protected by Cloud Security and Cloud Monitor free of charge.

1.Select mo	del 2.Complete configuration 3.Confirm configuration
Project	DEFAULT PROJECT
Security groups	New security group Existing security groups Operation guide
	Select a security group
	To open other ports, you can New security group E
Instance name	(Optional) Auto-create an instance name i Enter a name (up to 40 chars, 40 more allowed). For batch creation, this name will be taken
	Unique instance name 💿
Login methods	Set password SSH key pair Reset password after creation
Username	root
SSH key	Select a key Operation guide to
Security reinforcement	✓ Enable for free 安装组件免费开递DDoS防护 Details B 和主机安全基础版 Details B
Cloud monitoring	Enable for free FREE cloud monitoring, analysis, alarming, and server monitoring metrics (component installation required) Details E
Advanced settings	
Hostname	(Optional) Defaults to "Unnamed" if it's lef
	allowed.
Custom data	(Optional) The custom data is used to configure instances during launching. It supports the Shell format. The original data cannot exceed 16KB. The Shell script must start with "#" or "!" and the path to the interpreter reading the script (usually /bin/bash)
	The above input is encoded with base64.

2. After configuration confirmation and successful creation, you can view created launch configurations on the **Launch Configuration** page.



Valid 0 IT5.8XLARGE128 (32 core 128GB) Pay-as-you	45-you-go 5 Mbps System disk: SSD cloud disks Bill by traffic 50GB	

Step 2:Creating a Scaling Group

Last updated : 2024-01-08 17:53:29

Overview

A scaling group contains a collection of CVM instances that follow the same rules and have a shared purpose. This document introduces how to create a scaling group in the Auto Scaling console.

Directions

Creating a scaling group

1. Log in to the Auto Scaling console and click Scaling group in the left sidebar.

2. On the Scaling groups management page, click Create.

3. On the **Create scaling group** page that pops up, enter the basic information of the scaling group (fields marked with * are required).

Create scaling gro	up	×
1 Basic configu	uration > 2 Load Balancer > 3 Instance Allocation >	
4 Other config	gurations	
Name *	Please enter the name The name can contain up to 55 characters, including Chinese characters, English letters, numbers, underscores, hyphens and periods.	
Project	Default project	
Min capacity *	- 0 + 3	
Initial capacity *	- o + (i)	
Max capacity *	- 1 + ⁽	
Launch configuration *	The current launch configuration has only one mode. We recommend configuring multiple similar models to reduce the risk of scale-out failures. Configure Now 🖸	
Supported network *	If you don't have an available network, you can create a VPC 	
Support subnet *	Subnet ID Subnet name Availability zone	
	ne 2	
	You can select multiple subnets. CVMs will be created in these subnets randomly when auto-scaling up is triggered, so as to implement cross-subnet disaster recovery. Suggested settings 🛛	
	Next	

Name: Set the name of the scaling group.

Min capacity: The minimum number of instances allowed in a scaling group.

Initial capacity: The number of instances when the scaling group is created.

Max capacity: The maximum number of instances allowed in a scaling group.

Note:

The current number of CVM instances in the scaling group will be maintained between the minimum and the maximum scaling group capacities.

Launch configuration: Specifies the launch configuration to scale out CVM instances.



Supported networks and availability zones: Select networks and availability zones based on your needs.

4. Click Next.

5. (Optional) In the **Load balancer configuration**, associate an existing load balancing policy or create a CLB, and click **Next step: Instance allocation**.

Create scaling gro	up							×
Basic config	uration)	2	Load Balancer Configuration	>	3 Instanc	e Allocation	>	
4 Other config	gurations							
Load balancing 🛈	Please select	a cloud loa	ad balar 🔻 🗘					
	Instances creat existing LB or c If you need to o	ed in scalin reate one . configure n	g out will be mounted to the Learn More 🗹 nultiple load balancers, pleas	e associated lo e modify the s	ad balancer auton scaling group after	natically. You can s r creation.	select an	

6. (Optional) In Instance allocation, configure the spot instance allocation policy. You can also click Next step:Instance allocation to skip this step.

Note:

Only when you specify that the billing mode of the launch configuration is pay-as-you-go, you can create a scaling group with both pay-as-you-go and spot instances.

Enable **Spot instance allocation**.

Create scaling group		×
Basic configuration	> V Load Balancer > 3 Instance Allocation Configuration	>
(4) Other configurations		
Spot Instance Allocation		
Pay-as-you-go Base Capacity	- 0 + ⁽¹⁾	
Pay-as-you-go Above Base	- 70 + % (i)	
Spot instance creation policy	Capacity optimized 🔻 🛈	
Capacity rebalancing	Not Enable v 🚯	
Spot Fallback to Pay-as-you-go	Not Enable 🔻 🛈	
	Back Next: Other configurations Completed	

Pay-as-you-go base capacity: The minimum number of required pay-as-you-go instances in a scaling group. These instances have higher priority in auto scaling.

Pay-as-you-go above base: Controls the percentage of pay-as-you-go instances for the additional capacity beyond the base capacity. Valid range: 0-100

Spot instance creation policy: The policies for spot instance creation in a multiple-model lunch configuration.

Capacity optimized: First selects the most available spot instance model to make the best use of spot instance resources.

Lowest price: First launches spot instances at the lowest core price among the specified availability zones to minimize costs.

Capacity rebalancing: Enable it to replace a spot instance before it's terminated, thus maintaining the capacity and pay-as-you-go percentage of the scaling group.

Spot fallback to pay-as-you-go: Enable it to create pay-as-you-go instances when spot instances of the configured models are out of stock.

7. In the "Other configurations" step, refer to the following information to set the removal policy and instance creation policy.

Removal policy: When the scaling group wants to reduce the number of instances and has multiple choices, it determines which instances to remove according to the removal policy. The options **Remove the oldest instance** and **Remove the latest instance** are supported.

Instance creation policy:

Preferred availability zones (subnets) first: Based on the sequence of availability zones (subnets) configured, the system first selects configuration items higher in the sequence. If a failure occurs, the system automatically retries in sequence. This mode is suitable for scenarios with one primary availability zone and other secondary availability zones.

Multiple availability zones (subnets) distribution: During scale-out, the system will select availability zones (subnets) with relatively few instances in which to create new instances. This mode is suitable for architectures where instances need to be evenly distributed.

8. Click **Completed**. You can view the created scaling groups on the **Scaling groups** page.



Adding instances (optional)

- 1. Go to the **Scaling group** page and select the ID of the target scaling group to enter its details page.
- 2. Select the Associated to tab, click Add instances.

Scaling group	Scaling group details	Alarm trigger policy	Scheduled action No	tification Scaling activity Li
⊒‡ Launch Configuration	Add instances Remove s	elected		
	Instance ID/name	Monitor status	Life cycle	Removal protecti
	Total items: 0			

3. In the **Add instances** window that pops up, select the instance to be bound and click **OK**.

Note:

If you cannot add or remove a CVM instance to or from the list, check the maximum and minimum capacity values specified for the scaling group.

Step 3:Creating a Scaling Policy

Last updated : 2024-01-08 17:53:29

Overview

You can use scaling policies to increase or decrease the number of CVM instances in your scaling group: Create a **scheduled action** to perform scheduled scaling, which can be set to run periodically. Create an **alarm-triggered policy** to perform scaling based on Cloud Monitor metrics (such as CPU utilization and memory usage).

Directions

Creating a scheduled action

If your load changes are predictable, you can set scheduled actions to plan your scaling activities. This feature can automatically increase or decrease CVM instances according to a schedule. This allows you to flexibly cope with traffic load changes and improve device utilization while reducing deployment and instance costs.

1. Go to the **Scaling group** page and select the ID of the target scaling group to enter its details page.

2. Select the **Scheduled action** tab, and click **Create**.

← test							
Scaling group details	Associated to	Alarm trigger policy	Scheduled action	Notification	Scaling activity	Lifecycle Hook	
Create							
Name	Desc	ription					

3. In the **Create scheduled action** window that pops up, specify the action name, scaling group activities, repeat cycle, and other information.

4. After completing the configuration, click **OK** to view the scheduled action.

P	CICOL	
	Name	Description
	test	From 2022-06-27 16:00:00, change the Min capacity to 0, desired capacity to 0 and max capacity to 1. Execute once every 1 day(s). End time: 2023-06-27 16:00:00

Creating an alarm policy

If you want to perform scaling based on CVM metrics, you can create an alarm policy to plan device scaling. This policy helps you automatically increase or decrease the number of instances in your scaling group to flexibly handle business load changes, increase device utilization, and reduce deployment and instance costs.

Note:

When a scaling group is created, a ping unreachable alarm-triggered policy is created by default, which is used to replace unhealthy CVMs.

Before using an alarm policy, you need to install the latest version of Cloud Monitor Agent in the CVM image. For more information, see Installing Monitor Components.

- 1. Go to the **Scaling group** page and select the ID of the target scaling group to enter its details page.
- 2. Select the Alarm-triggered policy tab, and click Create.

Scaling group details	Associated to	Alarm trigger policy	Scheduled action	Notification	Scaling activity	Lifecycle Ho
Create	L		J			
Name		Description				
Total items: 0						

3. In the **Create alarm policy** window that pops up, configure the Cloud Monitor metrics (such as CPU utilization, memory usage, and bandwidth), which will be used as the basis of adding or removing CVM instances by a specified number or percentage.

You can also choose **Use existing policy (optional)** to copy an existing policy from an existing scaling group to the current scaling group.

Create alarm policy		×
Name *	Supports Chinese characters, English letters, numbers, underscores, hyphen	
Use Existing Policy (Optional)	Please select a scaling group Please select Copy	
if *	Instances in the scaling group:	
	CPU utilization 🔻 1 minute 🔍 Max 🔍 V V % Consecutive 1 V	
	Detailed Statistics Rules 😢	
Scaling group activities *	Increase 🔻 instances 🔻 cooldown second(s) (i)	
	OK Cancel	

4. After completing the configuration, click $\boldsymbol{\mathsf{OK}}$ to view the alarm-triggered policy.

	Create	
	Name	Description
	test	When the Max of CPU utilization is larger than 25 % in 1 min(s) for 1 consecutive times, the number of instances increase 20 CVM(s). The cooldown period is 30 seconds.
	Total items: 1	
Ľ		